One-Sided Switching Games

Objective
To describe how a defender can operate a system to optimally obfuscate its status from a strategic attacker, prolonging time to a successful attack.

One-Sided Switching Games

- Two agents: controller (defender), observer (attacker)
- At each time step, the controller decides to keep the system in the current state, or switch the system to a new state.
  - In the two-state case, the controller’s (mixed) strategy is $\lambda = (\lambda_1, \lambda_2) \in \Delta = [0,1]^2$, where $\lambda_i$ denotes the probability of staying in the state $i$.
  - A noisy observation (public signal) is generated probabilistically according to the underlying state.
- Furthermore, observer is aware when a state is revisited.
  - Uncertainty is thus over the true state labeling.
- Observer decides whether to: end the game by making a guess of the labeling (receives reward $r$ if correct, 0 otherwise) or wait for another observation (and discounting rewards by $\beta \in (0,1)$)

Belief state of the game:
The state of the game is the observer’s belief over the true state labeling. The belief is updated recursively, as expressed by $\pi' = f_\lambda(\pi, a, y')$, given the current belief $\pi$, the controller’s action $a$, the public signal $y'$, and the controller’s strategy $\lambda$.

Note: Due to the dependence of the observer’s belief on the controller’s strategy, the observer cannot update the belief state.

Game Properties
- For a fixed discount factor $\beta \in (0,1)$ the game has a value, which is characterized by the equation below.

The game’s value, denoted by $V(\pi)$, is the solution to the fixed point equation:

$$V(\pi) = \min_{\lambda \in \Delta} \max\{\pi r, (1 - \pi)r, \beta \mathbb{E}_{a,y'}[V(f_\lambda(\pi, a, y'))]\}$$

Notes:
- Controller only needs to reason about the observer’s belief, not the observer’s moves.
- The fixed point equation only yields the value of the game and the strategy of the controller (informed player) not the strategy of the observer (uninformed player).

Martingale beliefs:
The sequence of beliefs form a martingale for every control strategy $\lambda$. Furthermore, beliefs converge to a unique limit $\pi^*$.

By considering the state of the game as the probability distribution over the true state labeling, the game becomes a repeated game with the addition of a stop action (chosen by the observer)
  - Controller is switching the underlying state, but does not influence the true labeling of the states.
- Since beliefs are bounded, $\pi \in \Pi = [0,1]$, the sequence of beliefs converges by the monotone convergence theorem.

Convexity of expected continuation value:
The mapping $\pi \mapsto \mathbb{E}_{a,y'}[V(f_\lambda(\pi, a, y'))]$ is convex for every $\lambda \in \Delta$.

The martingale property of the belief state ensures that convexity of the value function translates into convexity of the expected continuation value.
- Critical property for efficient algorithm design.

Computation
- Define the value operator:

$$[TV]^*(\pi) = \min_{\lambda \in \Delta} \max\{\pi r, (1 - \pi)r, \beta \mathbb{E}_{a,y'}[V(f_\lambda(\pi, a, y'))]\}$$

Contraction property:
The operator $T$ is a contraction mapping with a unique fixed point, $V^*(\pi)$, representing the value of the game.
- Show: 1) $V(\pi) \leq W(\pi) \Rightarrow [TV]^*(\pi) \leq [TW]^*(\pi)$
- $2) \beta \in (0,1)$ s.t. $[TV + c]^*(\pi) \leq [TV]^*(\pi) + \beta c, c \geq 0$

Algorithm outline: The algorithm is a modified version of value iteration that samples both the belief space, $\Pi = [0,1]$, and the controller’s strategy space, $\Lambda = [0,1]^2$.
- Iterates monotonically converge to optimal: $\hat{V}^k(\pi) \rightarrow V^*(\pi)$
- The optimal strategy of the controller, $\lambda^*(\pi) = (\lambda^*_1(\pi), \lambda^*_2(\pi))$, is a greedy optimization under $V^*(\pi)$:

$$\lambda^*(\pi) = \arg\min_{\lambda \in \Delta} \max\{\pi r, (1 - \pi)r, \beta \mathbb{E}_{a,y'}[V^*(f_\lambda(\pi, a, y'))]\}$$

Form of value function:

Conclusions & Future Work
- Developed a new class of games and characterized strategies for optimally masking the status of a system from an adversarial observer with noisy measurements.
- Future work includes deriving an analytical solution of the fixed point equation and applying the theory to the design of robust learning systems.

AUTHORS:
Erik Miehling, Roy Dong, Cedric Langbort, Tamer Başar

POINT OF CONTACT:
Erik Miehling (miehling@illinois.edu)